This is the authors' final, peer reviewed manuscript published in
Empirical Economics 38(3) (2010) 0377-7332 with the title:
Is cannabis a gateway to hard drugs?
http://link.springer.com/article/10.1007%2Fs00181-009-0280-z

Is cannabis a gateway to hard drugs?

Hans Olav Melberg[1*], Andrew M. Jones[2] and Anne Line Bretteville-Jensen [1]

[1] Norwegian Institute for Alcohol and Drug Research, Oslo, Norway
[2] Department of Economics, University of Bergen, Bergen, Norway

[*] Norwegian Institute for Alcohol and Drug Research, Oslo, Norway. E-mail: hom@sirus.no

Is cannabis a gateway to hard drugs?

1 Introduction

Cannabis is sometimes said to be a gateway drug that increases the user's probability of taking up harder drugs like amphetamine, cocaine or heroin. The empirical basis for the hypothesis is the common finding that most hard drug users have started with less dangerous drugs first and that there seems to be a "staircase" from alcohol and solvents via cannabis and tablets to amphetamine, cocaine and heroin (see e.g. Kandel 1975). Although controversial, the hypothesis has had considerable influence on drug policy and legislation in many countries and has been a powerful argument in debates about legalization or decriminalization of cannabis. For instance, when asked about marijuana laws at a meeting in 2007, US Senator John McCain answered that "I believe that marijuana is a gateway drug. That is my view and that's the view of the federal drug czar and other experts, although that is also a debatable question"[1] In contrast to this, a former cabinet minister in the UK, Peter Lilley, has argued that the gateway argument "is the reverse of the truth. Making the supply and possession of cannabis a criminal offence drives people through the gates of the law into the illegal world in which they must acquire their supplies from people who may also push heroin, cocaine and other hard drugs."[2]

[1] The comment is publicly available on a video on the Internet (http://www.youtube.com/watch?v=JzkREVOVYsw) and is from meeting at Milton, New Hampshire on August 11, 2007.

[2] The comment was made in a debate in the House of Commons and is recorded in the Commons Hansard Debates text for Friday 9 Nov 2001, Volume No. 374, Part No. 048: Column 519.

Liberalizing cannabis laws may increase the number of users as it will lower the opportunity cost of using the drug and probably increase the physical and cultural availability (see e.g. Williams (2004) for a review of cannabis participation studies). Whether such liberalization subsequently will lead to an increase in the number of hard drug users depends, among other things, on whether there is a gateway effect or not. The purpose of this paper is to test whether there is evidence of such a causal link between the uptake of cannabis and hard drugs.

There are different pathways that might be the basis for a causal gateway effect in drug use. For instance Pudney (2003, p. 183c) lists three:

> "One is the nature of consumer preferences: the consumption of soft drugs may create a psychological or physiological need for further, stronger experiences of the same type. Another is through social interaction: the act of obtaining and using soft drugs may bring the user into contact with hard-drug users or suppliers whom they would not otherwise have met. A third is through information and credibility: experience of the use of soft drugs with no obvious ill effects may appear to contradict and undermine the strong negative publicity directed against the use of illicit drugs in general, so that advice against hard drugs becomes less persuasive."

In addition there may be an *adjustment cost effect*: for some individuals consuming any illicit drug may cross a psychological threshold that makes it less costly to proceed into another stage of drug use.

Testing for a gateway effect with retrospective survey data poses a statistical challenge as it involves dealing with selection bias due to unobserved individual heterogeneity. The apparent causal relationship in the observed association may be spurious since unobserved factors may influence both the probability of cannabis use and the use of other drugs. For instance, a traumatic childhood may be causally important for both cannabis and later heroin use. If this factor is unobserved, the effect of childhood trauma will be picked up by the variable for cannabis use. It would be

misleading to conclude that cannabis is a gateway drug since it picks up the effects of an omitted third variable.

Two recent tests of the hypothesis report that the gateway effect of cannabis is greatly reduced after taking unobserved heterogeneity into account, even to the point of not being statistically significant (Beenstock and Rahav 2002; Pudney 2003). Other recent studies claim, on the other hand, that although the gateway effect is reduced when unobserved heterogeneity is taken account of, there is still a significant association (van Ours 2003; Fergusson et al. 2006; Bretteville-Jensen et al. 2008). Hence, the core question of whether the observed sequential pattern of drug initiation is due to correlation or causality remains unresolved.

This paper builds on this literature and makes new methodological and empirical contributions. The paper develops a new approach for testing the hypothesis using a latent class bivariate hazard model in which both the intercept and all slope coefficients are estimated separately for each latent class. The main difference between this approach and the ones employed by van Ours (2003) and Pudney (2003), is that the individual heterogeneity is allowed to affect the marginal effects of all the variables and not only the random intercept. In this sense, it imposes fewer restrictions on the model that is used to test the gateway hypothesis.

Previous studies have attempted to identify the gateway effect by comparing patterns of initiation of hard drugs between cannabis users and non-users, effectively using non-users as the control group. This approach is problematic when there is unobserved heterogeneity that makes users and non-users systematically different from each other: comparing treatment and control groups may fail to compare like-with-like. In contrast, our strategy is to restrict the analysis to those who have used cannabis at

some point in their lives and to exploit variation in the timing of cannabis and hard drug initiation to identify the gateway effect. This avoids the problem of having a non-comparable control group. There may still be systematic unobservable heterogeneity within the group of cannabis users, but this is dealt with by using a bivariate hazard specification with shared frailty (see Abbring and van den Berg, 2003).

The empirical results benefit from a unique set of data on drug prices collected through face-to-face interviews with people visiting the needle exchange service in the city of Oslo, Norway. Data on drug prices are rare, especially price information provided directly by users. Changes in the relative prices of drugs may affect the choice of which drug to use first (if at all) and for this reason having reliable information about prices is important. Previous contributions have either adopted approaches in which price variables were ignored or have tried to create proxies for prices.

Our findings demonstrate, first of all, that there is a gateway effect and the hazard of taking up hard drugs increases substantially after the initiation of cannabis. Results from the univariate hazard model suggest that the hazard of starting to use hard drugs more than doubles (2.6) after controlling for the influence of other observable factors. Secondly, the results demonstrate the importance of taking unobserved heterogeneity into account. The effect of recent cannabis use differs considerably between two groups of users when a latent class analysis is employed. For a small group that we label “troubled youths”, we find that recent use of cannabis doubles the hazard of starting to use hard drugs. For the second group – labeled “most youths” - recent cannabis use makes less difference to the hazard of using hard drugs and the coefficient is statistically significant only at the 10% level. Although the relative hazard rate increases by 34 per cent, one should keep in mind that the marginal hazard for "most

youths" is very low (0.00023). Thus, the analysis indicates that the overall gateway effect is created by a large effect in a small group and a weaker effect for most people.

2 Identifying the gateway effect

To identify a possible causal effect of cannabis on subsequent use of hard drugs one needs to account for self-selection into the group of cannabis users and into the group of hard drug users. Ideally, to single out the effect of cannabis use, one would like to know the counterfactual outcome for cannabis users: what the probability of hard drug use would have been if they had not started to use cannabis in the first place. In practice this counterfactual cannot be observed and, in the absence of randomized experiments, attention must focus on alternative estimation strategies for handling the potential endogeneity of the gateway variable.

One response to the problem of spurious correlation between the use of cannabis and hard drugs is to adopt a selection on observables strategy and include as many as possible of the potential confounders in the analysis as control variables. Yamaguchi & Kandel (1984) and Ferguson & Horwood (2000) are two studies that have included a wide range of variables that are assumed to influence drug use and deviant behavior. A problem with this approach is that many variables affecting drug use are simply unavailable or are very difficult to measure. For instance, in addition to childhood traumas, time preferences are sometimes argued to be an important causal factor in the decision to use illegal substances (Ainslie 1992). Although it is possible to measure (indicators of) time preference, this is seldom done in large surveys of drug use in the

population. Similarly, studies of twins suggest that genetic factors are important in determining an individual's use of illegal drugs (Lyons et al. 1997), but this is, for all practical purposes, an unobserved variable for researchers who use general questionnaire-based surveys. In short, there are good reasons to expect that unobservable factors like emotional experiences, time preferences and genetics are important in the decision to use drugs and therefore it is necessary to take these unobservables into account when testing the gateway hypothesis that cannabis is a stepping stone to harder drugs.

One well-known approach is the instrumental variable (IV) technique. For instance, Pacula (1998) uses past prices of alcohol as instruments for previous consumption of the drug and estimates a gateway effect of alcohol on current use of marijuana. Other examples include DeSimone (1998), who uses two measures of state-level penalties for marijuana possession and two alcohol related variables (beer tax and parent's alcohol problems) as instrumental variables for predicting marijuana consumption and Beenstock and Rahav (2002), who apply variants of the IV approach to sequences of events using prices by birth cohorts as instruments. The main problem with the IV approach in general is finding valid instruments. From an economic perspective, prices can provide relevant instruments and alcohol and cigarette prices have been used frequently. These prices vary over time and between countries and states, but they cannot reflect contemporaneous individual differences in behavior within the same area. This does not imply that the instrumental variable approach should be rejected, but it seems worthwhile to explore alternative approaches. Our access to unique data on drug prices is used as part of our identification strategy but we

also adopt a method that does not necessarily rely on exclusion restrictions and hence the need for instruments.

Depending on the types of data at hand and assumptions one is willing to impose, there are various ways to take account of unobserved factors. It is frequently assumed that unobserved variables are stable over time, specific to the individual and influence a range of behaviors. When analyzing retrospective survey data using duration models, one needs to create time-to-event data by taking advantage of questions that relate to events that have occurred at some earlier point in time. The structuring of the data as a panel is based on an underlying approach in which one uses bivariate or multivariate hazard models. Analysing two or more behaviors simultaneously allows the unobserved heterogeneity to be captured. Employing models of this kind usually requires exclusion restrictions (see e.g., Heckman and Honorè 1989) but an important paper by Abbring and van den Berg (2003) shows that sample variation in the timing of treatments and outcomes provides identifying information in hazard models without the need for such exclusion restrictions. Heckman and Navarro (2007) extend Abbring and van den Berg's identification results in the context of dynamic discrete choice models for treatment effects.

The intuition of this approach is that particular individuals may have unobserved characteristics ("frailty") that make them more prone to both soft and hard drug use. This creates a potential selection bias in the sense that their hazard of initiating both soft and hard drugs will be higher, but without any direct gateway effect. However this frailty is modeled as time invariant heterogeneity using a shared frailty specification. Conditional on frailty, the treatment effect of interest (in our case the gateway effect) is identified by assuming that the actual timing of the treatment (initiation of cannabis use)

is random and is unaffected by the anticipation of the subsequent outcome (initiation of hard drugs). Under these assumptions the model is identified without exclusion restrictions. Here, as the assumptions may not be satisfied we also use exclusion restrictions, on drug prices, to strengthen identification.

Both van Ours (2003) and Pudney (2003) employ retrospective survey data and refer to Abbring and van den Berg's (2003) result as a source of identification in their hazard model specifications. van Ours (2003) analyses the use of cannabis and cocaine by employing a finite mixture approach to model the unobserved heterogeneity. Pudney (2003) takes a broader approach and analyses six categories of deviant behavior simultaneously, including minor offending and serious crime in addition to solvent abuse, soft drug use, social drug use and hard drug use. He employs a discrete time hazard model and a parametric specification of the unobserved heterogeneity which is estimated using maximum simulated likelihood estimation.

Given the nature of the information on drug use available in the Oslo survey data and the unique set of price data for illicit drugs at hand we too have chosen a hazard model approach. This means that identification of the causal gateway effect could be based solely on the timing of events, as demonstrated by Abbring and van den Berg. Compared to Pudney (2003) and van Ours (2003) we impose fewer restrictions on the model by allowing the unobserved variables to influence all the parameters of the model, not just the intercepts. Identification is reinforced by using the price data as instruments.

Our focus is on testing whether cannabis is a gateway drug and because of the overlap in use and the nature of the drugs, we have chosen to merge use of amphetamine, cocaine and heroin into one cluster for hard drugs; we test whether recent

cannabis use increases the probability of starting to consume a hard drug irrespective of which type of hard drug that might be.

2.1 The model

Our starting point is a bivariate mixed proportional hazard model with shared frailty of the type discussed by Abbring and van den Berg (2003) in their approach to the identification of treatment effects with duration data. In our application the treatment is use of the gateway drug and the outcome of interest is initiation of hard drug use. The first equation defines the hazard for initiation of cannabis use at a certain age, while the second equation estimates the hazard for initiation of hard drug use:

$$h_{it}^{1} = \lambda_1(t)\phi_1(x)u_1$$
$$h_{it}^{2} = \lambda_2(t)\phi_2(x)\delta(t \mid s,x)u_2$$

h_{it}^{j} is the hazard of individual (i) starting to use drug j (1=cannabis, 2=hard drugs) at age t, given the values of the covariates (x) and the shared frailty associated with an individual that affects both cannabis and hard drug use (unobservables u_1 *and* u_2 are correlated). $\lambda_1(.)$ and $\lambda_2(.)$ are the baseline hazards for cannabis and hard drug use. The function $\delta()$ captures the treatment effect, which depends on elapsed time (t), treatment time (s) and potentially on (t-s) the time between treatment and outcome. This formulation also allows for interactions with the covariates (x). This emphasis on the use of the timing of events to identify the treatment effects is exemplified by van Ours'

(2003) use of the concept of an "incubation period" to identify the gateway effect. He defines an indicator variable for whether use of cannabis had been initiated within a given number of years (the incubation period). We adopt a similar approach, using four years as the incubation period in our baseline model.

The observed covariates (*x*) used in this paper can be categorized in three groups. First, to isolate the effect of cannabis, we control for time-dependent covariates (the price of cannabis, amphetamine, cocaine and heroin). Second, given the correlation between gender, childhood problems and drug use, we have included gender and self-reported measures of serious childhood problems with parents, friends, police and school. Third, duration dependence is modelled using a cubic function of time (t). An alternative approach would be to use time dummies in a piecewise linear specification. But, in the interest of making our results more directly comparable to Pudney (2003), we chose to follow his specification and use a smooth function for duration dependence.

Our sample consists of individuals all of whom are initially at risk of starting both cannabis and hard drugs and these individuals are then followed over time and we adopt a discrete-time hazard specification (see e.g., Lancaster 1990; Jenkins 1995). To estimate the hazard models we reorganize the data so that each individual (1,…, n) is associated with multiple observations – one at each point in time from the initial period until either the individual starts to use the drug or the time of the survey interview if they are a right censored observation (t_i=1…t_{ij}). A new binary variable y_{it} is created that equals 1 for the period at which drug use begins and 0 otherwise. For those individuals who do not start within the survey period y_{it} always equals 0. For those who start, y_{it} only equals 1 in the final period and subsequently the individual is dropped from the

sample for all remaining periods. Using this reorganization of the data, the log-likelihood for a univariate discrete-time hazard function can be written as:

$$\log L = \sum_{i=1}^{n} \sum_{t=1}^{t_{ij}} y_{it} \log[h_{it}] + \sum_{i=1}^{n} \sum_{t=1}^{t_{ij}} (1 - y_{it}) \log[1 - h_{it}]$$

This log-likelihood takes the form of a standard binary choice model applied to the expanded dataset. Common choices of functional form for h_{it} are the complementary log-log model, which is the discrete-time equivalent of a continuous-time proportional hazard specification, and the logit model, which gives a non-proportional hazard specification.

This discrete-time specification can now be extended to the bivariate model for both cannabis and hard drug use. The sample likelihood function for the bivariate model is:

$$L = \prod_{i=1}^{n} E_u \left\{ \prod_{j=1}^{2} \left\{ \prod_{t=1}^{\tau_j - c_j} \left[1 - h_{it}^{j} \right] \right\} h_{it}^{j\,c_j} \right\}$$

where c_j is a dummy indicating a non-censored case. For each hazard function, the product runs from t=1 to t=τ-1 for uncensored observations and t=1 to t=τ for censored. An observation is censored if substance j hasn't been initiated by the end of the survey. The presence of unobservable heterogeneity (u) means that the two hazard functions must be estimated jointly and $E_u\{\}$ denotes the expectation over the joint distribution of unobserved heterogeneity. Pudney (2003) adopts a parametric approach by assuming

that the u's are jointly normally distributed and using maximum simulated likelihood estimation to deal with the numerical integration. van Ours (2003) adopts a semiparametric approach using a bivariate finite density estimator for u.

2.2 Latent class specification

In contrast to earlier studies, we adopt a latent class approach to deal with the presence of unobserved heterogeneity in the hazard functions. This approach has several advantages over previously applied methods for dealing with unobserved heterogeneity in models of the gateway effect. One advantage is that we can relax the parametric assumptions that are necessary in other approaches that rely on Gauss-hermite quadrature or maximum simulated likelihood estimation. We do not make assumptions about the parametric distribution of the unobserved heterogeneity. Another advantage is that the approach imposes fewer restrictions on the parameters of the model. For instance, in Pudney (2003) and van Ours (2003), only the constant term in the regression is assumed to differ depending on whether or nor the person has a high or low value on the unobserved characteristic. This is restrictive because there is no a priori reason to assume that the unobserved heterogeneity does not affect the marginal effects of the other variables. A person with adverse childhood experiences, for instance, may react differently to changes in prices than a person with a normal childhood. Hence, this more general approach allows all coefficients to vary.

The latent class model assumes that each individual is drawn from one of K possible sub-groups or latent classes that exist in the population where π_k indicates the

share of the population that belongs to the group *k*.[3] The parameters of the hazard function for each drug may be different depending on which group the individual belongs to.

Given that the individual belongs to one of the two (or more) groups, and since we do not know which, we may use π_k to denote the probability that individual *i* is a member of group k. The probability of an observed pattern for drug use is then the sum of the probability of observing the pattern conditional on group membership, weighted by the respective probabilities of being in each group. In this case the log-likelihood function for the whole sample can be written as:

$$LogL = \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K} \pi_k \left\{ \sum_{j=1}^{2} \left\{ \sum_{t=1}^{\tau_j - c_j} \log\left[1 - h_{it}^{jk}\right] + c_j \log\left[h_{it}^{jk}\right] \right\} \right\} \right\}$$

where

$$0 \le \pi_k \le 1 \forall k, \sum_{k=1}^{K} \pi_k = 1$$

This approach to capturing unobserved heterogeneity helps us solve the problem of maximizing the likelihood function because it reframes the maximization problem in a way that allows us to use the EM algorithm (see for example, Ng et al. 2002).[4] This is

[3] Previous results indicate that relatively few classes are needed (see for example Deb and Trivedi 1997)

[4] The expectation (E) step uses Bayes' rule for posterior probabilities to give an updated and improved estimate of the probability that an individual is a member of each latent class and these estimates can be substituted for the unknowns in the likelihood function. When we have estimates for the unknown individual heterogeneity (interpreted as probabilities of being a member of a group) it is possible to use weighted estimation of the binary choice models to estimate the coefficients of the hazard functions using that particular set of probabilities of group membership (this is the maximization (M) step). These coefficients give rise to a new set of estimates of the contributions to the likelihood which, in turn, can be used to improve the

possible since the unobserved heterogeneity is integrated out of the likelihood, in this case by factoring the sum over latent classes. The probabilities of class membership are treated as parameters to be estimated.

3 The data

We combine two different data sources. The first dataset was collected through postal questionnaires sent to a representative sample of 21-26 year olds living in Oslo in 2002. It provides information on the development of drug use initiation in a general population of youths. The data do not, however, contain any price information on illicit drugs so a second data source is required. Since 1993 the Norwegian Institute for Alcohol and Drug Research (SIRUS) has on a regular basis conducted face-to-face interviews with drug addicts visiting the needle exchange service in Oslo and prices for various types and quantities of drugs have been recorded

3.1 Data for drug use

The response rate for the postal questionnaires was roughly 50 percent, with more women than men answering the questions (see Table 1). A reminder was sent and a total of 1,984 questionnaires were registered. The respondents reported their experience with licit and illicit drugs in addition to information on age, gender and possible childhood

---

estimates of class membership, and so on, until the likelihood converges. The algorithm is implemented in a Stata v.9 program that is available from the authors on request.

problems with parents, friends, school and police. More than 40 per cent of the sample reported to have tried cannabis at least once in their life time. As mentioned, in order to reduce the problem of unobserved heterogeneity and compare like-with-like we have confined our analyses to these cannabis users only. For comparison, however, Table 1 displays descriptive statistics for the cannabis users as well as for the sample as a whole.

(Table 1 about here)

There was no age difference between the full sample and the smaller group of cannabis users, the average age was 24 years in both. The proportion of males, however, was higher among the cannabis users (44 per cent versus 38 per cent) and a larger fraction of the cannabis users reported childhood problems with parents, school, friends or police (22 per cent versus 16 per cent stated at least on of these). Problems with parents were reported most frequently in both samples (13 and 9 per cent, respectively) whereas problems with the police were stated by 3 per cent of the cannabis users and 2 per cent of the whole sample.

As expected, the cannabis users reported a higher lifetime prevalence of alcohol and illicit drugs. Practically every cannabis user had used alcohol (99 per cent) compared to 92 per cent in the full sample, and the corresponding numbers for amphetamine and cocaine use were 30 versus 13 per cent and 26 versus 11 per cent, respectively. The heroin prevalence was relatively low in both groups (3 and 1 per cent, respectively. For all drugs, substantially smaller fractions reported frequent use, defined as use on 25 or more occasions. The debut ages did not differ between the groups, and in line with the pattern for initiation found in other studies, the average debut ages

suggest that drug consumers in Oslo start to use alcohol prior to cannabis, then proceed to amphetamine, heroin and cocaine.

There are also significant overlaps between the use of different drugs. The low prevalence of heroin implies that many in the sample have used only amphetamine or cocaine without also using heroine. However, almost all heroin users had also used other drugs (amphetamine and or cocaine), and 61% of those who had tried either amphetamine or cocaine had also used the other drug.

There is no way of knowing which came first for individuals who report that they have used both cannabis and hard drugs in the same year (6% of the final sample). In order not to make the unsupported assumption that cannabis came first, our specification makes the probability of using hard drugs depend on cannabis consumption in the previous years, not the current year.

Figure 1 illustrates the "staircase" in drug use initiation where the highest hazard rate for starting with alcohol peaks at an earlier age than the highest hazard rate for cannabis use and use of amphetamine and cocaine. The hazard rate figure for heroin is left out due to the small sample size (n=24) and compared to the other substances it indicates a less uniform pattern. The hazard rates give the probabilities for various age groups of starting with a drug given that the person has not started up to that age.

(Figure 1 about here)

The "staircase" pattern is also confirmed when we examine the individual ages of initiation for the various drugs. 92% of heavy drug users claimed to have used cannabis before any of the hard drugs. Less than 5 per cent reported a lower debut age

for amphetamine than for cannabis, 1 per cent a lower debut age for cocaine and no one claimed to have started with heroin before they tried cannabis for the first time. Further, nearly all hard drug users seem to have used cannabis at some point and only 4 of the 254 amphetamine users, 5 of the 218 cocaine users and 1 of the 24 heroin users claimed no cannabis experience.

It is well known that in general surveys like the one used here, homeless and institutionalized people are under-represented, as are people with various sorts of deviant behavior. However, the relatively high reported prevalence of illicit drug use in the present sample indicates that many drug users do respond to postal questionnaires. Still, to validate our data we have compared prevalence rates for life-time drug use to two recent Norwegian studies employing alternative sampling procedures but covering the same age groups. The Norwegian Institute for Alcohol and Drug Research conducts on a regular basis national surveys in which interviewers visit households drawn from a central register (see Horverak (2006) for more details about the procedure). Alcohol related questions are asked face-to-face whereas a self-filled questionnaire is used for information about illicit drug use. In addition a longitudinal dataset based on a representative sample of youngsters recruited in a 1992 school survey (mean age 15) and followed to 2005 (mean age 28) collected drug use information (Pedersen, 2008). Despite different survey designs, these datasets yield comparable prevalence figures to the present. All three samples, however, are likely to be more representative for recreational than hard core drug users.

Recall bias may be a problem, as people are asked to recall the debut age of behaviors in the past. One may argue, however, that using an illicit drug for the first time is a unique event and that users will tend to remember it. In line with this, one

recent study of response reliability in adolescent substance use progression suggests that the initiation sequences were reported consistently when checked again three years after the first interview (Golub et al. 2000).

3.2 The price data

In order to examine the economic aspects of drug use more than 3,500 interviews with drug injectors visiting the needle exchange service in Oslo have been conducted since 1993. Interview sessions were first held on a monthly basis, then quarterly from June 1994 and bi-annual since September 1997. The interviews were anonymous. In addition to some background variables, interviewees were asked detailed questions about their level and source of monthly income, levels of drug consumption and the prices they had paid for the different types and quantities of drugs. Given both the number of interviews and the fact that this was the only major needle-exchange service in Oslo during this period, the data is considered to accurately reflect the overall price level of drugs in Oslo. More details about the sampling procedure and representativeness can be found in Bretteville-Jensen and Biørn (2003).

We use annual median prices for the different drugs and the nominal prices are deflated by the consumer price index (CPI 1998=100). The prices are reported for small quantity sales: we use the price of 1 gram of cannabis, ¼ gram of amphetamine and cocaine and 1/24 gram of heroin. The survey does not provide price data for the period 1988-1992. Such data are needed for a minority of the sample since some of the youngsters in the sample set turned 12 during those years (The starting age for inclusion

in the pseudo-panel data set). According to information from the drug section of the Oslo police nominal drug prices were very stable during the first period, so we have used deflated 1993-prices for those years. For cocaine, police records of prices are used throughout the period.

With the exception of cocaine, inflation adjusted drug prices were substantially reduced the last half of the study period (Figure 2). Heroin in particular, has become cheaper. In 1988 users were charged more than 400 Norwegian kroner (NOK) for the smallest unit sold at the market whereas the equivalent price in 2002 was less than 100 NOK. The price of amphetamine has also fallen substantially whereas cocaine users paid around the same in 2002 as in 1988. Nominally, the price of one gram of cannabis has been stable throughout the period although the CPI deflated price has fallen.

(Figure 2 about here)

4 Results

We first present the results from the univariate and the bivariate hazard models, then a more detailed examination of the gateway variable. Thereafter, we examine more closely the two groups that the model identifies and test the robustness of the model. As our main interest is the possible gateway effect of cannabis, only the results for the hazards of hard drug use are presented (the full set of results are available upon request). Our preferred specification uses the logistic hazard model and relies on exclusion restrictions (the price variables) as well as on the timing of events to identify the gateway effect.

The univariate hazard model for initiation of hard drugs provides a benchmark for the subsequent analyses. The results from running a separate logistic hazard equation, that does not take unobserved heterogeneity into account, suggest that cannabis is a statistically significant stepping stone to the use of harder drugs (see Table 2). The dummy variable for cannabis initiation in the past four years is statistically significant at the 1-percent level. Childhood problems and the price of heroin also seem to have a significant impact on the uptake of hard drug use and the hazard increases with time (age).

(Table 2 about here)

The results from the bivariate hazard model with shared frailty reveal that the picture changes when unobserved heterogeneity is taken into account. In the latent class model each individual is assumed to belong to one out of two groups. For those belonging to Group 1, the coefficient for recent cannabis use is statistically significant only at a 10 per cent level. In contrast, cannabis is a statistically significant predictor of later hard drug use for those who belong to Group 2. Among the other explanatory variables, Table 2 shows that the coefficients for childhood problems, heroin price and time are statistically significant for Group 1 whereas gender, childhood problems, amphetamine and cocaine prices and time are statistically significant for Group 2.

(Figure 3 about here)

Group membership is unknown but posterior probabilities of class membership can be computed using Bayes rule. A histogram of these posterior probabilities reveals

that most of the individuals share a high probability of having a latent characteristic that makes them belong together in one group (labeled "Group 1", see Figure 3) To learn more about the groups, we split the sample so that those with an estimated probability of 0.5 or more for belonging to Group 1 are assigned to that group (n=617) whereas the remaining individuals are put in Group 2 (n=194). The descriptive statistics of individuals in Group 1 show no extreme values on any of the background variables (see Table 3). The second group is interesting in the sense that it consists of a subset of what one might call "troubled youths." They do worse on the background variables (childhood problems with police, school, friends and parents) as well as reporting to have started their illicit drug use at an earlier age (e.g. mean debut age for cannabis is 15.6 versus 18.9 in Group 1). Group 2 not only states a much higher life time prevalence of amphetamine, cocaine and heroin use but also a more frequent use of illicit drugs. More than 80 per cent among the "troubled youths" reported that they had used cannabis on more than 25 occasions and 57 per cent had as frequent use of amphetamine. The corresponding numbers for the larger group of sample are 35 and 29 per cent, respectively.

(Table 3 about here)

Figure 4 and Table 4 provides more details about the hazard rates and the magnitude of the estimated gateway effect for the two groups. Figure 4 compares the average estimated hazard for the use of heavy drugs in Group 1 and Group 2 at different ages. Both curves peak at around the age of 20, but the estimated hazard is much larger for Group 2 than Group 1. This difference underlines the argument that it is important to

distinguish between groups when discussing the gateway effect since the magnitude of the effects seems to differ greatly depending on which latent class someone belongs to.

(Figure 4 about here)

In Table 4 the first row presents estimates of the hazard rate for the 'untreated' (h^0) evaluated with recent initiation of cannabis (the gateway effect) set to zero and at the means of the other regressors. This gives a sense of the rate of initiation of hard drug use in the absence of a gateway effect. As expected the estimated probability of initiating use of hard drug without recent cannabis use is low and close to zero for Group 1. All estimates for the hazard rate increase when evaluated with recent use of cannabis set to one, as shown in the second row. The third row presents estimates of the partial effect ($h^1 - h^0$): the difference in hazard rates with and without recent use of cannabis evaluated as the mean of the regressors. This shows how the hazard of hard drug initiation is increased by the gateway effect. In absolute terms recent cannabis use is estimated to have a relatively small impact on the initiation of hard drugs for Group 1 (0.00023) and a larger one for Group 2 (0.025). As the last row shows, the odds ratios for the gateway effect (h^1 / h^0) are substantial. For Group 2 this is statistically significant and recent use of cannabis almost doubles the hazard of later hard drug use. One should, however, be careful not to equate a statistically significant gateway coefficient with a strong gateway effect. Moreover, the results do not imply that cannabis users in this sample are twice as likely to become *frequent* hard drug users/addicts since only a small minority of those who try hard drugs report to have used any hard drug 25 times or more.

(Table 4 about here)

An interpretation of the main findings is that for most people, the influence of recent cannabis use is relatively low and only statistically significant at a 10 per cent level. Still, from this level, the risk of taking up drugs like cocaine, amphetamine or heroin increases by one third after initiating cannabis. On the other hand, the sample consists of a minority of individuals whose hazards for hard drugs do increase significantly after having used cannabis. The unobserved characteristic – genetic composition, time preferences or upbringing - seem to make the individuals in this group more vulnerable to the influence cannabis may have on later hard drug use.

5 Checks for robustness

Computation of latent class models is prone to problems due to local optima. To examine the robustness of the results, the analysis was repeated with different starting values (for individual group membership) and the process of convergence was traced. The algorithm produced a likelihood that increased gradually and monotonically towards a maximum. Repeated tests using different starting points also showed that the routine converged to the same solution regardless of the starting point, indicating that the maximum really is a global maximum.

In an effort to further assess the sensitivity of the results to the specification of the model and the data, we used the same latent class model but with changes in the model specification. The aim was to examine the stability of the results. Clearly one

would not expect precisely the same estimates, but if the general pattern is robust one would be more inclined to accept the conclusions and less fearful that the estimates were driven by specification issues in the modelling.

First, the model also gave robust results when we shortened the incubation period: the length of time in which cannabis could be viewed as causally important (Table 5). Decreasing the original 4-year time window to 2 years and re-running the analysis produced comparable results: two distinct groups, one small group with a statistically significant gateway coefficient and another group in which the effect was smaller and/or statistically insignificant (at the 5-percent level).

Second, models with no exclusion restrictions were tested – both by including and excluding the price of all the drugs in all the equations. Both specifications gave similar general results: a small group for whom cannabis was an important gateway drug and a larger group in which it was weak or statistically not significant. This result is worth emphasizing since previous studies have not had access to the similar data on prices. Given that the price of heroin fell during the period of our analysis, one would expect an increase in the uptake of hard drugs. A model which did not include prices might attribute this general trend to a larger gateway effect. Excluding prices could then lead to an overestimation of the effect. In line with this we find that the coefficient on the gateway dummy is larger when we exclude price information, but at the same time we also show that including prices does not eliminate the gateway effect for the "troubled youths." In this way the results seem to support the overall conclusion that this is a vulnerable group.

Third, an alternative specification with different variables was tested. In this model cocaine price was left out both because its relevance was low (cocaine

prevalence in Norway was not very high at that point in time) and the data on cocaine prices was from a different source than the other price data (police records instead of our own interviews). Also the problem variable was split so "problems with the police" became a separate variable as opposed to being a part of the aggregate index that was originally used as a "problem variable." This model also gave results that were in line with the benchmark model which further strengthens the belief that the general result is not driven by minor issues in the specification.

In the empirical analysis presented it is assumed that the unobserved heterogeneity divides the sample into two groups. To test whether this is a reasonable assumption, a model with three groups was also estimated. However, in this model the solutions became unstable as the algorithm gave different answers for different starting values. This reflects the problems of local optima, identification and convergence for over-parameterized models. For more on accuracy and problems of estimating these kinds of models see Gaure et al. (2007).

Finally, in order to reduce unobserved heterogeneity the benchmark model was originally estimated using only those respondents who had tried cannabis (41% of the full sample). To test the importance of this restriction, we estimated a model using the full sample. Once again the algorithm separated the sample into one large and one smaller group (34% of the sample). The marginal effect of recent cannabis use on the probability of using hard drugs was statistically significant in the small group at the 10% level, but not in the large group. Also the marginal effect was almost three times larger in the small group than the larger group (0.0037 vs. 0.0013). This indicates that the general direction of the results are robust both to changes in the specification and to changes in the sample.

(Table 5 about here)

6 Conclusions

The commonly observed sequential pattern of drug use initiation may well be explained by an increased risk of starting to consume a more harmful drug after first having used a soft drug but there could also be other factors influencing the uptake of various drugs. As no survey, no matter how detailed, will include all potentially important variables, testing the influential gateway hypothesis imposes a statistical challenge. This paper has argued that the issue of unobserved individual heterogeneity is central, that the empirical methods should take into account time-varying covariates like prices and that it should make as few restrictions on the model as possible. After developing such a model, based on a bivariate discrete time hazard model with shared frailty, the conclusion from the empirical findings is that the gateway effect of cannabis is statistically significant for a sub-group of "troubled youths." Within this group the risk of taking up hard drugs doubles after the initiation of cannabis. For most youths, however, the gateway coefficient was only statistical significant at a 10% level, and the marginal effect was relatively small. Still, the results indicate that the hazard for hard drug initiation would increase by one third during a four year "incubation period" following cannabis initiation.

The results underline the importance of adopting a general approach in which all the coefficients are allowed to vary. The large differences between some of the coefficients in the two groups indicate this and statistical tests of the hypothesis that the coefficients in the two latent classes are similar are clearly rejected. This may explain

some of the discrepancy between our results and Pudney (2003) and van Ours (2003). Pudney, and to some extent van Ours, argue that the gateway effect is greatly reduced after taking unobserved heterogeneity into account. A more general test in which all variables are allowed to vary reaches a more nuanced conclusion in which the gateway effect actually increases for almost a quarter of the sample when comparing their results to that of the full sample.

The finding of a statistical significant gateway effect is, however, in line with recent studies using alternative approaches for taking account of unobserved heterogeneity. DeSimone (1998), employing an IV technique, finds that previous cannabis use increases the probability of subsequent cocaine use by 29 percentage points. As mentioned in Section 2, finding valid instruments are challenging. Fergusson et al. (2006) examined the association between *frequency of cannbis use* and the use of other illicit drugs. They report a strong association between cannabis use and subsequent hard drug use and it was particularly strong during adolescence. Their chosen fixed effect approach, which requires panel data with sufficient within-variation in the dependent variable and regressors, cannot be used however, to model "one-off" decisions such as whether or not to try a harder drug. Bretteville-Jensen et al. (2008) examine the influence of previous drug use on subsequent *regular* use (>25 times) of cannabis, amphetamine and cocaine. Although substantially reduced after unobserved factors are taken into account they find a statistically significant gateway effect for all three drugs tested. Unfortunately, the multivariate probit approach cannot take time-varying covariates, like drug prices, into account. The present study adds to this gateway literature by introducing a flexible approach which, among other things, allows for testing separate gateway effects for subgroups of cannabis users.

As previously described, there are at least four possible mechanisms underpinning an observed gateway effect. A better understanding of these mechanisms is needed but beyond the scope of this paper. It is important to notice, however, that the finding of a gateway effect need not necessarily undermine moves to liberalize cannabis laws. If the driving force behind cannabis users' initiation of hard drugs is that they come in contact with hard drug users whom they not otherwise would have met, a separation of the markets for soft and hard drug, for instance in line with the Dutch model, could be an option. Further, if people, after experiencing no obvious ill effects of soft drug use, have reduced confidence also in the strong negative publicity directed against hard drug use, the solution may be to make the differences between the various drugs more distinct in information material and campaigns directed against illicit drug use. If, on the other hand, it is the addiction or the adjustment cost effects that operate, a liberalization of penal sanctions may be undesirable, even though the number of additional hard drug users resulting from the policy change may be quite modest.

Whatever mechanism or combinations of mechanisms that operates it seems clear that one group of cannabis users are more vulnerable to the effect cannabis use has on subsequent hard drug use. We saw that not only did they start to consume illicit drugs at a younger age, a substantially larger proportion also used drugs like amphetamine and heroin more intensively, i.e. more people in this group reported to have used these drugs on more than 25 occasions. They also reported more childhood problems with police, parents, friends and school. Whether the underlying reasons are childhood traumas, negative peer influence, high time preferences, less beneficial genetic endowments or other factors, the "troubled youths" seem to have less resistance with respect to further drug involvement. Thus, some policy implications of the findings

may still be suggested: Early identification of the “troubled youths” and adequate help to minimize the effect of their risk factors can be effective in reducing the number of hard drug users.

Acknowledgements The authors would like to thank the editor and referees for helpful comments.

References

Abbring JH, van den Berg G (2003) The Non-Parametric Identification of Treatment Effects in Duration Models. Econometrica 71:1491-517

Ainslie G (1995) Picoeconomics. Cambridge University Press, Cambridge

Beenstock M, Rahav G (2002) Testing Gateway Theory: Do Cigarette Prices Affect Illicit Drug Use? J Health Econ 21: 679-98

Bretteville-Jensen AL, Biørn E (2003) Heroin Consumption, Prices and Addiction: Evidence from Self-Reported Panel Data. Scand J Econ 105:661-79

Bretteville-Jensen AL, Melberg HO, Jones AM (2008) Sequential Patterns of Drug Use Initiation – Can We Believe In the Gateway Theory? BE J Econ Analysis Policy 8 (Contributions), Article 1

Deb P, Trivedi PK (1997) Demand for Medical Care by the Elderly: A Finite Mixture Approach. J Appl Econ 12:313-36

DeSimone J (1998) Is Marijuana a Gateway Drug? Eastern Econ J 24:149-64

Fergusson, DM, Horwood LJ (2000) Does Cannabis Use Encourage Other Forms of Illicit Drug Use? Addiction 95:505-20

Fergusson DM, Boden JM, Horwood LJ (2006) Cannabis use and other illicit drug use: testing the cannabis gateway hypothesis. Addiction 101:556-569

Gaure S, Roed K, Zhang T (2007) Time and causality: A Monte Carlo assessment of the timing-of-events approach, J of Econometrics 141:1159-1195

Golub A, LaBouvie E, Johnson BD (2000) Response Reliability and the Study of Adolescent Substance Use Progression. J of Drug Issues 30:103-18

Heckman JJ, Honorè BE (1989) The identifiability of the competing risks model. Biometrica 76:325-330

Heckman JJ, Navarro S (2007) Dynamic discrete choice and dynamic treatment effects. J Econometrics 136:341-396

Horverak Ø (2006) Norwegian drinking pattern – undergoing change? In: Bryhni A (ed) Alcohol and Drugs in Norway, Annual report, Norwegian Institute for Alcohol and Drug, Oslo, pp. 28-39

Jenkins S (1995) Easy Estimation Methods for Discrete-Time Duration Models, Oxford Bull Econ Statist 57:129-38

Kandel DB (1975) Stages in Adolescent Involvement in Drug Use. Science 190:912-14

Lancaster, T (1990) The Econometric Analysis of Transition Data, Cambridge University Press, Cambridge

Lyons MJ, Toomey R, Meyer JM, Green AI, Eisen SA, Goldberg J, True WR, Tsuang MT (1997) How Do Genes Influence Marijuana Use? The Role of Subjective Effects. Addiction 92:409-17

Ng SK, Krishnan T, McLachlan GJ (1995) The EM Algorithm. In: Gentle JE, Härdle W, Mori Y (eds.) Handbook of Computational Statistics, Springer, pp. 137-68

Pacula RL (1998) Does Increasing Beer Tax Reduce Marijuana Consumption? J Health Econ 17:557-85

Pedersen W (2008) Hasjbruk blant unge (Young people's cannabis use) The Norwegian Medical Journal 128(16): 1825-1828

Pudney S (2003) The Road to Ruin? Sequences of Initiation to Drug Use and Crime in Britain. Econ J 113:182-98

van Ours JC (2003) Is Cannabis a Stepping-Stone for Cocaine? J Health Econ 22:539-54

Williams J (2004) The Effect of Price and Policy on Marijuana Use: What Can Be Learned From the Australian Experience? Health Econ 13:123-137

Yamaguchi K, Kandel DB (1984) Patterns of Drug Use from Adolescence to Young Adulthood: III. Predictors of Progression. Am J Public Health 74:673-81

Table 1 Variable definitions and descriptive statistics (Full sample, n=1,984; Cannabis users only, n= 811)

Variable Label	*Variable Definition*	*Full sample*		*Cannabis users*	
		Mean	Std.dev.	Mean	Std.dev.
Age	Age in years	24.1	1.64	24.0	1.66
Gender	Dummy; 1 if male	0.38	0.49	0.44	0.50
Parents	Dummy; 1 if problems with parents	0.09	0.29	0.13	0.34
School	Dummy; 1 if problems at school	0.07	0.26	0.10	0.30
Friends	Dummy; 1 if problems with friends	0.04	0.20	0.05	0.21
Police	Dummy; 1 if problems with police	0.02	0.13	0.03	0.16
Childhood Problems	Dummy; 1 if any problems	0.16	0.37	0.22	0.41
Alcohol	Dummy; 1 if ever used alcohol	0.92	0.27	0.99	0.09
Cannabis	Dummy; 1 if ever used cannabis	0.41	0.49	1	1
Amphetamine	Dummy; 1 if ever used amphetamine	0.13	0.33	0.30	0.46
Cocaine	Dummy; 1 if ever used cocaine	0.11	0.31	0.26	0.44
Heroin	Dummy; 1 if ever used heroin	0.01	0.11	0.03	0.16
Hard drugs	Dummy; 1 if ever used hard drugs	0.16	0.36	0.36	0.48
Alco-age	Debut age for first use of alcohol	15.3	2.25	14.5	1.96
Cann-age	Debut age for first use of cannabis	18.0	2.76	18.1	2.73
Amph-age	Debut age for first use of amphetamine	18.8	2.39	18.8	2.30
Hero-age	Debut age for first use of heroin	19.5	2.57	19.6	2.67
Coca-age	Debut age for first use of cocaine	20.0	2.32	19.9	2.32

Table 2 Coefficient estimates for hazard models for hard drug use

	Single equation model	Latent class models (taking unobserved heterogeneity into account)	
		Group 1	*Group 2*
Male	0.109	0.115	0.569***
	(0.122)	(0.161)	(0.161)
Childhood problems	0.448***	0.490***	0.627**
	(0.080)	(0.112)	(0.115)
Price of amphetamine	-0.002	0.000	-0.022***
	(0.003)	(0.003)	(0.004)
Price of cocaine	0.000	0.000	0.001**
	(0.000)	(0.000)	(0.000)
Price of heroin	-0.003***	-0.014***	0.002
	(0.001)	(0.003)	(0.002)
T (time)	1.876***	3.435***	2.182***
	(0.463)	(1.126)	(0.494)
t2	-0.181***	-0.285**	-0.201***
	(0.057)	(0.12)	(0.061)
t3	0.005**	0.007*	0.004*
	(0.002)	(0.004)	(0.002)
Previous use of cannabis	0.742***	0.294*	0.710***
	(0.128)	(0.163)	(0.185)
Constant	-8.310***	0.115***	-5.029***
	(1.303)	(0.161)	(1.449)

* Significant at 10-percent level

** Significant at 5-percent level

*** Significant at 1-percent level

Table 3 Differences between the clusters in the latent class analysis

	Group 1, n = 617 (Most youths),		Group 2, n = 194 ("Troubled youths"),	
Variable	*Mean*	*St. dev.*	*Mean*	*St. dev*
Gender (percentage male)	44.2	0.497	42.3	0.495
Percentage with reported use of …				
… amphetamine (0/1)	13.0	0.336	85.1	0.357
… heroin (0/1)	0.6	0.080	8.8	0.283
… cocaine (0/1)	11.0	0.313	72.7	0.447
Mean age of starting to use …				
…cannabis	18.9	2.518	15.6	1.485
…amphetamine	20.6	1.887	18.1	2.044
…heroin	21.0	1.547	19.3	2.845
Frequency of use (percentage reported to have used the drug more than 25 times) …				
…cannabis	35.2	0.478	82.5	0.381
…amphetamine	28.8	0.455	57.0	0.497
…heroin	0	0	41.2	0.507
Percentage with serious childhood problems				
… with the police	1.3	0.113	7.2	0.259
… in school	8.9	0.285	13.9	0.347
… with friends	4.5	0.208	5.2	0.222
… with parents	13.0	0.336	14.9	0.357

Table 4 Estimates of gateway effects for hazard models for hard drug use (standard errors in parenthesis)

	Single equation model	Latent class models Taking unobserved heterogeneity into account	
		Group 1	*Group 2*
Hazard without recent cannabis use: h^0	0.011*** (0.002)	0.00067* (0.00031)	0.026*** (0.006)
Hazard with recent cannabis use: h^1	0.027*** (0.005)	0.00091* (0.00046)	0.051*** (0.012)
Gateway effect on hazard rate: $h^1 - h^0$	0.016*** (0.004)	0.00023 (0.00018)	0.025*** (0.009)
Relative gateway effect: h^1 / h^0	2.584*** (0.343)	1.342* (0.218)	2.034*** (0.376)

* Significant at 10-percent level

** Significant at 5-percent level

*** Significant at 1-percent level

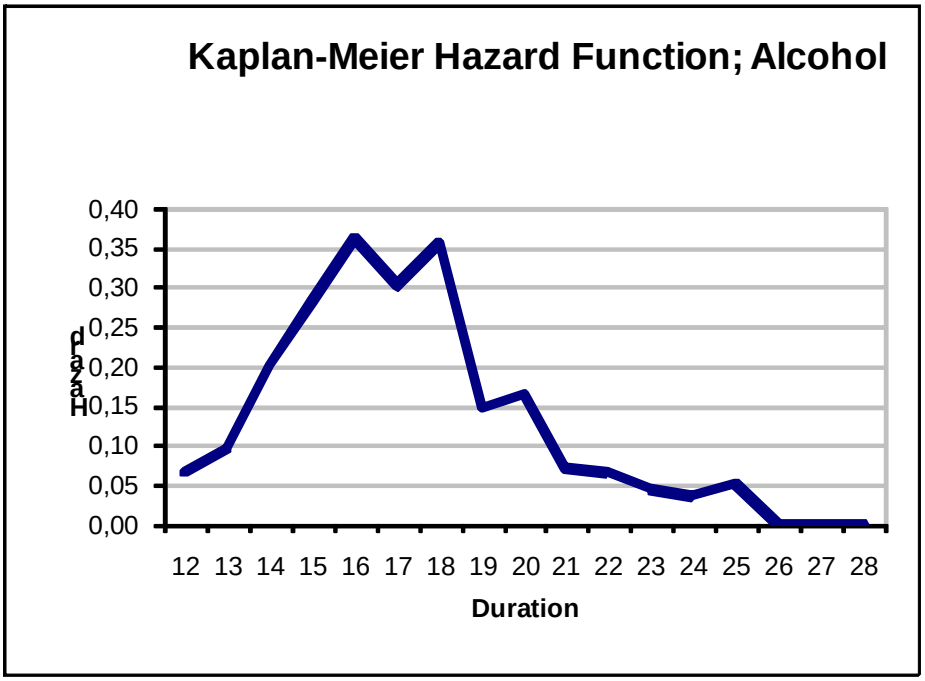
Table 5 Robustness analysis

Specification	Gateway coefficient, group I	Group size	Gateway coefficient, group II	Group size
Benchmark model	0.710***	194	0.295*	617
Short incubation period (2 years)	0.684***	162	0.351*	649
Including all price variables in all equations	0.53***	178	0.43	633
Including no price variables	0.93***	234	0.000	577
Using only amphetamine as hard drug	1.3***	216	-1.0***	595
Using only cocaine as hard drug	0.53***	223	26.4	588
New specification (see text)	0.93***	230	0.28	645

* Significant at 10-percent level

*** Significant at 1-percent level

Fig. 1 Hazard rates for the onset of alcohol, cannabis, amphetamine and cocaine use



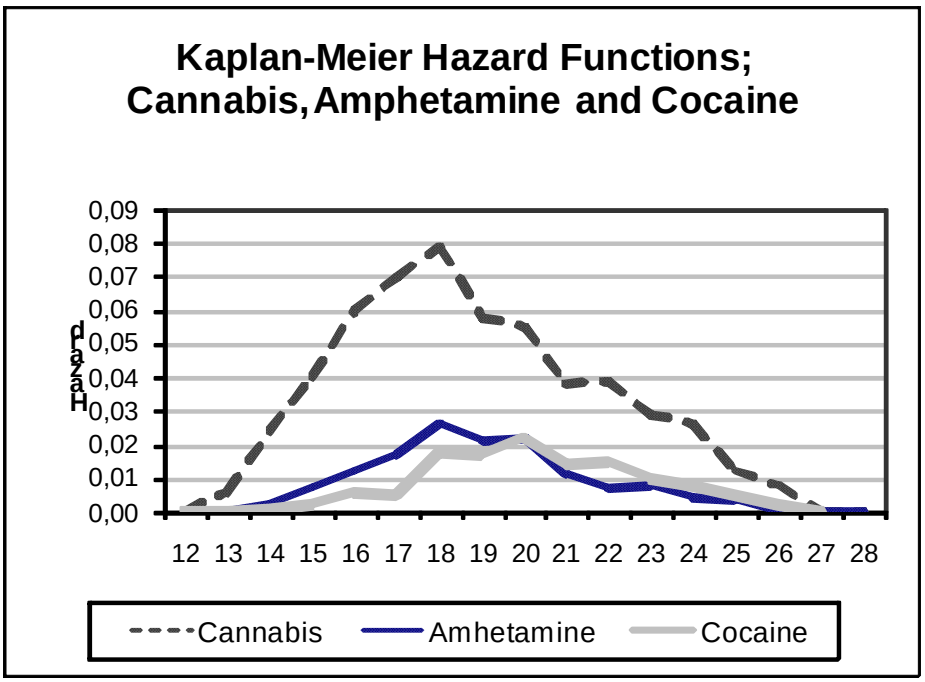
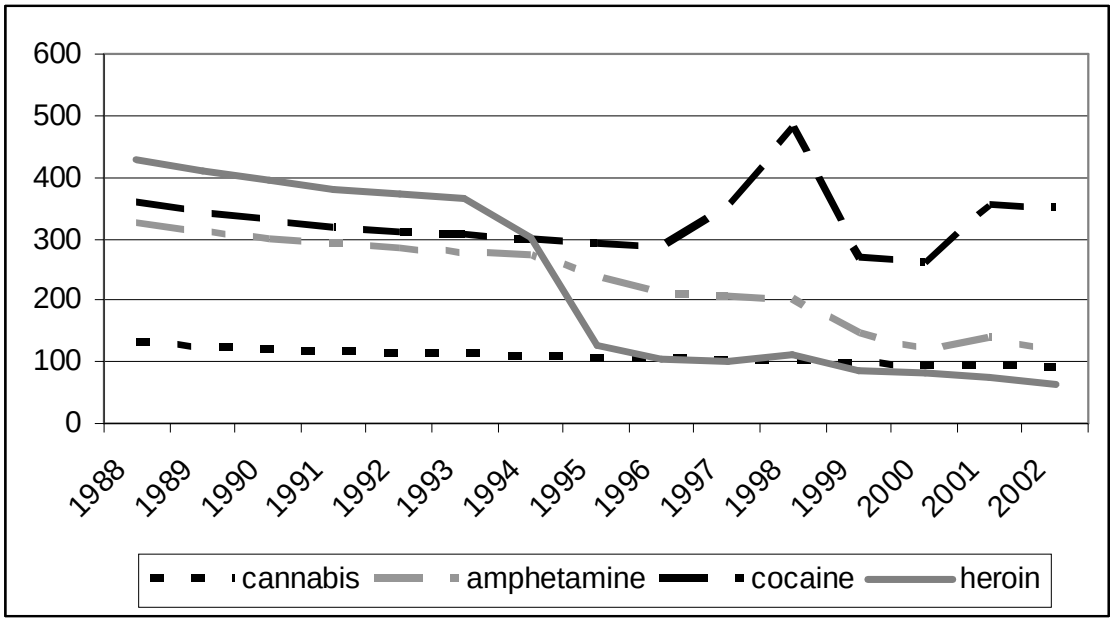


Fig 2 Prices of cannabis, amphetamine, cocaine and heroin, 1998-2002*



*Note: The prices relates to the following quantities of the drugs: 1 gram of cannabis, ¼ gram of amphetamine, ¼ gram of cocaine and 1/24 gram of heroin. The prices have been deflated by the CPI (1998=100).

Fig 3 Histogram of latent class membership probabilities

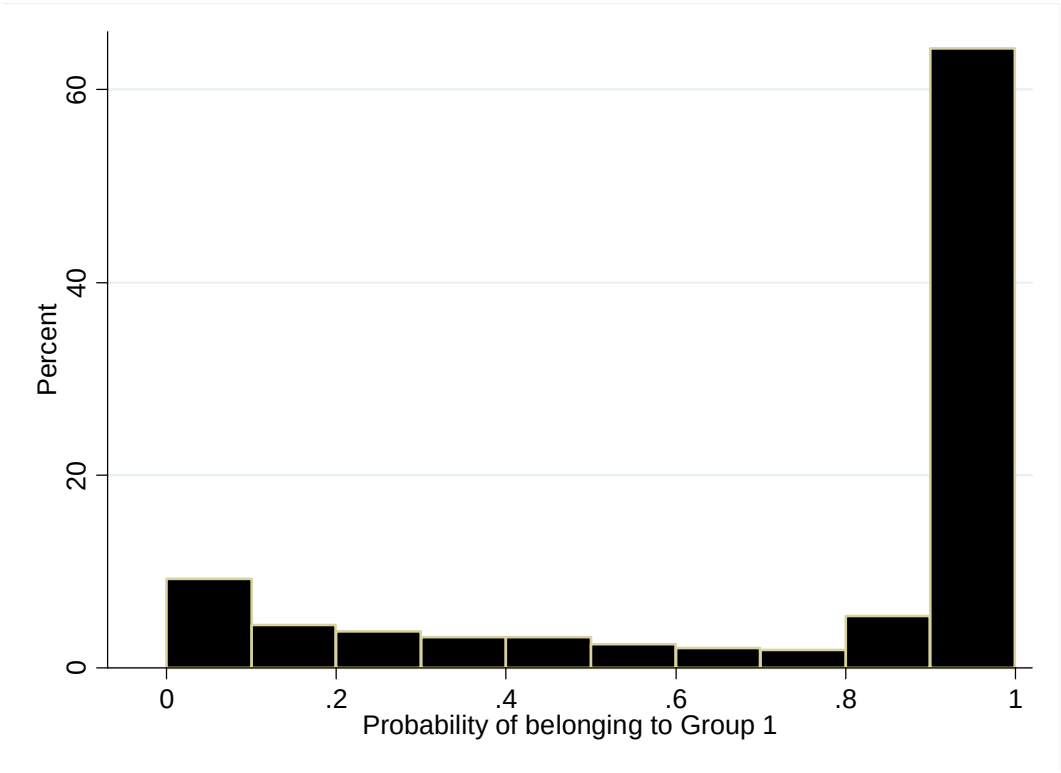


Fig 4 Estimated mean hazard rates for the use of hard drugs in Group 1 and Group 2